4.0 *IN VIVO* REFERENCE DATA USED FOR AN ASSESSMENT OF TEST METHOD ACCURACY

4.1 Description of Protocol Used to Generate *In Vivo* Data

4.1.1 Draize Rabbit Eye Test

The test method protocol most widely accepted by regulatory agencies for the evaluation of ocular eye irritants is based on the Draize rabbit eye test method. The methodology, originally described by Draize et al. (1944), involves instillation of 0.1 mL of the test substance (e.g., liquids, solutions, and ointments) into the conjunctival sac of an albino rabbit eye. Injuries to the cornea, conjunctiva, and the iris are examined and scored at selected time intervals after exposure. Scoring is subjective and based on a discrete, arbitrary scale (**Table 4-1)** that continues to be used today for grading the severity of ocular lesions. The scores for the observed ocular injuries range from 1 to 2 for iris effects, from 1 to 3 for conjunctival redness and discharge, and from 1 to 4 for corneal effects and conjunctival chemosis. A score of zero is assigned when the eye is normal and no adverse effects are observed. Injuries to the eye were originally observed up to four days after application of the test substance. However, in current practice, these time points vary according to the degree of irritation, the clearing time, and testing requirements imposed by the various regulatory agencies.

The original Draize protocol describes a scoring system in which each ocular parameter is graded on a continuous numerical scale. The scores may be weighted (as shown in **Table 4-1**); however, most classification systems today do not use such a weighting factor. The weighting of the score by Draize et al. (1944) is biased more heavily for corneal injury, since injury to the cornea has the greatest probability for producing irreparable damage to the eye. For example, the scores for the degree of corneal opacity (which range from 0 to 4) and the area of cornea involved (scored on a scale of 0 to 4) are evaluated for each animal. The values are then multiplied together and then by a factor of 5; the maximal corneal score is 80. The iris score is multiplied by a factor of 5 to give a maximal score of 10. The scores for the

Table 4-1 Scale of Weighted Scores for Grading the Severity of Ocular Lesions*

Lesion	Score**
Cornea	
A. Opacity – Degree of density (area which is most dense is taken for reading	
Scattered or diffuse area – details of iris clearly visible	1
Easily discernible translucent areas, details of iris slightly obscured	2
Opalescent areas, no details of iris visible, size of pupil barely discernible	3
Opaque, iris invisible	4
B. Area of cornea involved	
One quarter (or less) but not zero	1
Greater than one quarter but less than one-half	2
Greater than one-half but less than three quarters	3
Greater than three quarters up to whole area	4
Score equals A x B x 5 Total maximum = 80	
Iris	
A. Values	
Folds above normal, congestion, swelling, circumcorneal injection (any one or all of these or combination of any thereof), iris still reacting to light (sluggish reaction is positive)	1
No reaction to light, hemorrhage; gross destruction (any one or all of these)	2
Score equals A x 5 Total possible maximum = 10	
Conjunctiva	
A. Redness (refers to palpebral conjunctiva only)	
Vessels definitely injected above normal	1
More diffuse, deeper crimson red, individual vessels not easily discernible	2
Diffuse beefy red	3
B. Chemosis	
Any swelling above normal (includes nictitating membrane)	1
Obvious swelling with partial eversion of the lids	2
Swelling with lids about half closed	3
Swelling with lids about half closed to completely closed	4
C. Discharge	
Any amount different from normal (does not include small amount observed in inner canthus of normal animals	1
Discharge with moistening of the lids and hairs just adjacent to the lids	2
Discharge with moistening of the lids and considerable area around the eye	3
Score equals (A + B + C) x 2 Total maximum = 20	

The maximum total score is the sum of all scores obtained for the cornea, iris and conjunctiva.
* From Draize et al. (1944)
**Scores of 0 are assigned for each parameter if the cornea, iris, or conjunctiva are normal.

three conjunctival parameters are added together and then the total is multiplied by a factor of 2 for a total score of 20. The overall score for each animal is then developed by adding the calculated values with a maximum total score for eye irritation of 110.

4.1.2 Current *In Vivo* Ocular Irritation Test Method Protocols

Since the original description of the *in vivo* rabbit eye test method, regulatory agencies in the United States, as well as in other countries, have modified the test method protocol to suit their specific needs and goals in protecting human health (**Table 4-2**). Regulatory agencies generally recommend using healthy adult albino rabbits (e.g., White New Zealand). The eyes of each test animal are examined within 24 hours prior to test initiation. A quantity of 0.1 mL (for liquid test substances) or 0.1 g (for solid, granular, or particulate test substances) is placed into the conjunctival sac of one eye of each animal, after pulling the lower lid from the eyeball. The lids are held together for about one second to decrease loss of test substance from the eye. The other eye remains untreated. Although the observation period varies depending on the testing guideline or regulatory agency, the eyes are typically examined periodically at 24 hour intervals for at least 72 hours after application of the test substance for adverse effects to the cornea, conjunctiva, and iris. The length of the observation period should be sufficient to evaluate reversibility or irreversibility of any the observed effects, but generally does not exceed 21 days. The specific ocular effects observed are generally the same as those described by Draize et al. (1944) in **Table 4-1**, with the exception being that other lesions, such as pannus[1] and herniation of the cornea, are often noted. In addition, corneal, iris, and conjunctival lesions are scored using the individual numerical grades described in **Table 4-1**, but weighted scores and an overall score for irritation are not calculated or used for U.S. and European regulatory purposes.

Depending on the regulatory agency, the numbers of animals required for a study of ocular irritation can vary. To minimize pain and suffering of rabbits exposed to potentially corrosive agents, the EPA and European regulatory agencies suggest that if a test substance is anticipated to produce a severe effect (e.g., corrosive effect), a test in a single rabbit may be conducted. If a severe effect is observed in this animal, further testing does not need to be conducted. In cases where more than a single animal needs to be tested, at least three animals should be examined to classify the ocular effects produced by the test substance

[1] Pannus, also known as "chronic superficial keratitis", describes a specific type of corneal inflammation. Pannus is caused by a local inflammatory response that begins within the conjunctiva, and with time spreads to the cornea. On a cellular level, the inflammation is composed of brown melanin pigment, red blood vessels, and pink scar tissue.

67 **Table 4-2 Regulatory Guidelines for *In Vivo* Ocular Irritation Test Methods**

Test Method Component	Reference				
	Draize et al. (1944)	**OECD TG 405 (April 2002)**	**FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)**	**FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)**	**European Union Annex V B.5 (formerly EEC; EU 1992)**
Evaluate existing animal & human eye data	NA	Yes	Yes*	NS	Yes
Results from dermal irritation study	NA	Yes	Yes*	Yes	Yes
Perform SAR for eye irritation	NA	Yes	Yes*	NS	Yes
Screen for pH	NA	Yes	Yes*	Yes	Yes
Results from validated alternative ocular methods	NA	Yes	Yes*	Yes	Yes
Animal model/Number of animals					
Animal species and strain	Albino rabbit	Healthy young adult albino rabbits.	New Zealand White rabbit	Healthy adult albino rabbits recommended. Other mammalian species may be substituted with justification.	Healthy young adult albino rabbits.
Sex and weight	NS	NS	Sex NS; 2.0-3.0 kg	NS	NS
Screen for severe effects	NS	1 animal – further testing not required if substance produces corrosive or severe effects	NS	1 animal – further testing not required if substance produces corrosive or severe effects	A single animal test should be considered if marked effects are anticipated
Main test/ confirmatory test	NS	Up to 2 additional animals, with sequential testing if irreversible effects suspected. If 2nd animal has severe effects, test is discontinued. Additional animals may be used to confirm weak or moderate responses.	A minimum of 6 animals, and up to 18 animals for confirmatory tests.	≥ 3 animals	≥ 3 animals

Test Method Component	Reference				
	Draize et al. (1944)	**OECD TG 405 (April 2002)**	**FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)**	**FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)**	**European Union Annex V B.5 (formerly EEC; EU 1992)**
Test substance (amount and method of application)					
Liquids	0.1 mL	0.1 mL	0.1 mL	0.1 mL	0.1 mL
Solids, pastes, particulates	NS	0.1 mL, or ≤ 100 mg	0.1 mL, or ≤ 100 mg	0.1 mL, or ≤ 100 mg	0.1 mL or 0.1 g
Aerosols	NS	Single burst of about 1 second sprayed at 10 cm	NS	Single burst of about 1 second sprayed at 10 cm	NS
Pump sprays	NS		NS	0.1 mL	NS
Application of test substance	Test substance is placed in the conjunctival sac.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.	Test substance is placed in the conjunctival sac of one eye.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.
Use of anesthetics prior to instillation of test substance	NS	Local anesthetic may be used, if the test substance is anticipated to cause pain.	Local anesthetic may be used prior to instillation of test substance.	Local anesthetic may be used, if the test substance is anticipated to cause pain.	Anesthetic may be used after 24 hours if it does not influence response of the eye to irritants.
Observation					
Observation Period	At least 48 hours. Extended if irritation persists.	At least 72 hours, except when animal shows severe pain or distress, or early severe/ corrosive effects, upon which the animal is humanely killed. Otherwise, sufficient to evaluate reversibility or irreversibility within 21 days.	At least 72 hours. Extended if necessary.	At least 72 hours, but not more than 21 days. Should be sufficient enough to evaluate the reversibility or irreversibility of effects within a 21-day period.	At least 72 hours, except when animal shows severe pain or distress, or early severe/ corrosive effects, upon which the animal is humanely killed. Can be extended up to 21 days if effects persist.
Examination times after treatment	1, 24, 48 hours, and 4, 7 days.	1, 24, 48, 72 hours, 7, 14, 21 days.	24, 48, 72 hours, and 7 days	1, 24, 48, and 72 hours. Up to 21 days to assess reversibility.	1, 24, 48, and 72 hours. Can be extended up to 21 days.

Test Method Component	Reference				
	Draize et al. (1944)	**OECD TG 405 (April 2002)**	**FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)**	**FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)**	**European Union Annex V B.5 (formerly EEC; EU 1992)**
Observation aids	NS	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	
Irrigation					
Washout	NS	Generally, eyes may not be washed until after 24 hours post-treatment, except for solids, which may be removed with saline or water after 1 hour.	After 24 hours post-treatment, eyes may be washed with a sodium chloride solution.	After 24 hours post-treatment, eyes may be washed with water to show whether washing palliates or exacerbates irritation.	After 24 hours post-treatment, eyes may be washed.
Additional testing to determine effects of timely irrigation	NS	Not recommended unless scientifically justified.		Indicated when substances are shown to be irritating. At 30 seconds after exposure, the eyes are washed with water for 30 seconds	Possibility of washing out irritant at 30 seconds after treatment if previous evidence of severe effects.

Abbreviations: CPSC = U.S. Consumer Product Safety Commission, EEC = European Economic Community, EPA = U.S. Environmental Protection Agency, FDA = U.S. Food and Drug Administration, FIFRA = Federal Insecticide, Fungicide, and Rodenticide Act, NA = Not applicable, NS = Not specified, OECD = Organization for Economic Cooperation and Development, OPPTS = Office of Prevention, Pesticide, and Toxic Substances, OSHA = U.S. Occupational Safety and Health Administration, SAR = Structure activity relationships, TG = Test guideline, TSCA = Toxic Substances Control Act.
* Use of this information is not provided in the regulations cited, but in the CPSC Animal Testing Policy guideline (CPSC 1984) states that prior human experience, literature sources which record prior animal testing or limited human tests, and expert opinion may be used in making appropriate hazard determinations.

(EU 1992; EPA 1998). In contrast, regulations for other U.S. agencies (e.g., CPSC, FDA) require that at least six animals be examined to classify the effects produced by a test substance (CPSC 2003). The differences in current *in vivo* eye irritation test protocols in the United States appear to reflect the different objectives of eye irritation testing for industrial

chemicals regulated by the EPA versus the household consumer products, pharmaceuticals, cosmetics, and toiletries regulated by the CPSC and the FDA.

Various data transformations have been suggested as a means of comparing and rating irritants of varying severity. One of these measures is the MAS in which the Draize scores obtained at each time point are averaged across all animals in the study and the highest score obtained is taken as the MAS. This MAS value was later modified to the Modified Maximum Average Score (MMAS), which represents the highest average MAS value beginning with the 24 hour time point (ECETOC 1998).

4.1.3 Current *In Vivo* Ocular Irritancy Classification Systems

Although the *in vivo* eye irritation test method protocols are similar across different U.S. and international regulatory agencies, interpretation of the results from the *in vivo* test method varies considerably. As described in **Section 1.1**, several classification systems are in use for regulatory ocular irritancy testing purposes (**Table 1-2**). In the United States, two major classification systems are currently used, the FHSA guideline (FHSA 1964), which is used by the FDA, OSHA, and CPSC, and the 1996 guideline set forth by the EPA (EPA 1996). The FHSA guideline states that a test substance is considered an eye irritant if four or more of six rabbits have positive ocular scores in nonirrigated eyes within 72 hours after instillation of the test substance in the conjunctival sac (FHSA 1964). A positive score is defined by corneal opacity or iritis scores of $\geq$1, or conjunctival redness or chemosis scores of $\geq$2. In addition, if only one of the six rabbits shows ocular effects within 72 hours, the test substance in considered nonirritating to the eye. If two or three animals have positive ocular scores, the test is repeated in a second group of six rabbits. Then, if the criteria for an ocular irritant for the second test (three or more positive animals) or a nonirritant (0 positive animals) are met, a classification is made. However, if only one or two animals have positive scores in the second test, the test is repeated a third and final time. If one or more animals have positive ocular scores in the third test, the test substance is classified as an ocular irritant. If none of the animals have positive ocular scores in the third test, the test substance is classified as a nonirritant (FHSA 1964).

The EPA classification guideline (EPA 1996) takes into consideration the kinds of ocular effects produced in the *in vivo* rabbit eye test, as well as the reversibility and the severity of the effects. However, unlike the FSHA system, incidence is not considered, as classification is based on the animal that exhibits the most severe response in a group of three or more nonirrigated eyes. Data from all observation times are used for EPA classification. A positive score is defined by corneal opacity or iritis scores of ≥1, or conjunctival redness or chemosis scores of ≥2. EPA labeling regulations also require an assessment of the reversibility of positive scores. If a positive score persists for >21 days, the substance is classified as a Category I eye irritant, which is defined as "corrosive (irreversible destruction of ocular tissue) or corneal involvement or irritation persisting for >21 days". Substances that cause positive corneal opacity, iritis, or conjunctival scores that clear in 8-21 days are designated as Category II eye irritants. If positive scores induced by a substance clear within seven days, the substance is labeled Category III. Minimal effects (i.e., inconsequential or complete lack of irritation) produced by a test substance are designated as Category IV (EPA 1996).

In the current EU classification system for eye irritation, risk phrases are assigned to the label of a substance based on whether two or more of three animals exhibit a score, averaged across the 24-, 48- and 72-hours observation times for each ocular lesion, that falls within or above certain ranges of scores (**Table 1-2**) (EU 2001). Thus, the incidence and severity of effects are taken into consideration, but typically not the reversibility of effects. Hazard classification in the EU system corresponds to the following risk phrases: 1) R36 denotes "Irritating to eyes"; 2) R41 denotes "Risk of serious damage to the eyes". An *in vivo* rabbit eye study that results in a mean corneal opacity score ≥3 or a mean iris score of 2 in two or more of three animals would be assigned the R41 risk phrase. For studies in which six animals are used, the mean score for each ocular lesion for all animals in the study is calculated and used for classification and labeling purposes in the EU system. The criteria for assigning the risk phrase R36 are provided in detail in **Table 1-2**.

The GHS for the classification and labeling of hazardous chemicals (UN 2003) is an initiative developed through the cooperative efforts of the International Labour Office, the

OECD, and the UN to promote an internationally-harmonized approach for classifying chemicals according to their health hazards. For the purpose of harmonizing classification of ocular irritants, the GHS adopted an approach put forth by the OECD (1996) in its *Final Report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods*. A tiered testing and evaluation strategy using available data from dermal irritation studies, knowledge of structure activity relationships, pH screening, and data from validated alternative toxicological methods has been proposed (UN 2003). In addition, a single harmonized hazard category is proposed for irreversible effects on the eye/serious damage to eye (Category 1). Irreversible effects according to the GHS system include grade 4 cornea lesions at any time during the test, positive responses (e.g., iritis ≥ 1 or conjunctival chemosis ≥2) that do not reverse within 21 days, and cases where two or more of three animals exhibit a mean score (24, 48, 72 hours) for corneal opacity ≥3 and/or iritis >1.5. A single harmonized hazard category, Category 2, is proposed for reversible effects on the eye; however, for regulatory authorities that prefer to distinguish irritants in this group, subcategories have been developed based on whether effects reverse within 7 or 21 days. Category 2A is defined as an eye irritant with effects that fully reverse within 21 days. Category 2B is considered mildly irritating to the eyes, and is designated for substances whose effects reverse fully within seven days. Reversible effects include positive responses in two or more of three animals, where the mean score (24, 48, 72 hours) for corneal opacity or iritis ≥1, or conjunctival redness or chemosis ≥2.

4.2 Detailed Reference Data Used to Assess *In Vitro* Test Method Accuracy

The BCOP studies evaluated in this document include *in vivo* reference data generated using the basic procedures described above for the *in vivo* rabbit eye test method. For the Gautheron et al. (1994) study, the *in vivo* reference data were obtained from concurrent *in vivo* studies performed by Dr. J. Giroux at the Agence du Medicament in Montpelier, France. Studies were performed according to European Economic Committee (EEC) (1984 and 1991) guidelines with a few modifications. Three rabbits were used per test substance and MAS (Draize et al. 1944) were calculated. Only the MAS and day 1 scores for the 52 compounds are presented in the Gautheron et al. (1994) publication. The substances were

classified by the study authors according to both EEC (1984) and Kay and Calandra (1962) systems. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal, for 12 of these substances are available in the ECETOC Reference Chemicals data bank (ECETOC 1998). These 12 substances have been classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix E**).

For the EC/HO validation study (Balls et al. 1995), MMAS were calculated for the 59 test substances from existing and concurrently run *in vivo* studies, all of which were performed according to OECD TG 405 and following GLP guidelines. The data were generated since 1981 and met the following criteria:

- normally used at least 3 New Zealand White rabbits tested at the same time
- 0.1 mL or the equivalent weight of substance was instilled into the conjunctival sac
- anesthesia was not used
- observations were made at least at 1, 2, and 3 days after instillation.

The MMAS were developed from Draize scores calculated 24 hours or more after instillation of the test substance. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal, for each of these substances are available in the ECETOC Reference Chemicals data bank (ECETOC 1998). These substances have been classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix E**).

In the Swanson et al. (1995) study, *in vivo* reference data were obtained from standard (100 µL of test material; 7 formulations) or modified (30 µL of test material; 13 formulations) Draize eye irritancy tests. A MAS(30) or a MAS(100) is reported for each test substance. *In vivo* categories reported in the publication are mild (2 substances), mild/moderate (2), moderate (4), moderate/severe (1), severe/corrosive (4), and corrosive (7), and are based on an internal classification scheme used at S.C. Johnson & Son, Inc. Subsequent to the publication, the sponsor of the study, S.C. Johnson & Son, Inc., assigned GHS (UN 2003)

and EPA (1996) classifications to the substances and provided these classifications, along with detailed *in vivo* data for each test substance, to NICEATM. NICEATM verified these EPA and GHS ocular irritancy classifications, and also classified the test substances based on the EU (2001) ocular irritancy classification system (**Appendix E**).

For the CTFA Phase III study, data were obtained from a modified Draize eye test. Details of the protocol are provided in Gettings et al (1996). Six rabbits (three male, three female) were used for each test substance. The right eye of each rabbit was anesthetized prior to instillation of 0.1 mL of test substance into the conjunctival sac. Ocular irritation was evaluated at 1 hour, and at 1, 2, 3, 4 and 7 days. If irritation persisted, ocular responses were observed at seven day intervals up to a maximum of 21 days. MAS were determined according to Williams et al. (1982). Data were classified according to the scheme proposed by Kay and Calandra (1962) and the FHSA (1947). MAS, maximum average total scores for each endpoint (cornea, iris, conjunctiva), number of positive responses, maximum day to clear, and FHSA and Kay/Calandra irritancy categories are reported in the paper for the 25 test substances. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal, for each of these substances were provided by the CTFA. The substances have been classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix E**).

For the European Community prevalidation study (Southee 1998) of the BCOP assay, detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each animal, for each of these substances was available in the ECETOC Reference Chemicals data bank (ECETOC 1998). The substances have been classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix E**).

For the Casterton et al. (1996) study, the authors noted that they used *in vivo* reference data from existing sources. Fifteen of the test substances evaluated in the BCOP test method were selected from the formulations tested in the CTFA Evaluation of Alternatives Program – Phase III, and 48 were selected from the substances included in the ECETOC Eye Irritation

Reference Chemicals Data Bank (ECETOC 1992). Twenty-one test substances were Amway products with historical *in vivo* data, while the remaining substances were surfactant raw materials with *in vivo* data available from the suppliers. Only a subset of these data were available to NICEATM. The Access Business Group provided copies of original study reports containing *in vivo* reference data for 13 of the Amway product formulations evaluated in Casterton et al. (1996). Detailed *in vivo* data for the 15 surfactant-based formulations tested in Gettings et al. (1996) were available from the CTFA. *In vivo* data for 32 other substances were available in ECETOC (1998).

S.C. Johnson and Son, Inc. provided detailed *in vivo* reference data for eight of the 13 test substances evaluated in the Swanson and Harbell (2000) study of ethanol containing insect repellent formulations. The standard Draize eye irritancy test protocol was used for these eight test substances, utilizing six animals per substance.

ExxonMobil Biomedical Sciences, Inc. provided detailed *in vivo* reference data for the 16 petrochemical products evaluated by Bailey et al. (2004). All substances had been tested previously using the standard Draize eye irritancy test protocol, which consisted of instilling 0.1 mL of undiluted test substance into the conjunctival sac of three or six rabbits.

4.3 *In Vivo* Classification Criteria Used for Analysis

The *in vivo* rabbit eye database used to conduct a retrospective analyses of the accuracy of the BCOP test method includes studies that were conducted using from one to six animals. However, some of the *in vivo* classification systems considered for the accuracy analyses are currently devised to be applied to studies using no more than three animals. Thus, to maximize the amount of data used for the evaluation of BCOP, as well as for the three other *in vitro* test methods (ICE, IRE, HET-CAM) being evaluated, the decision criteria for each classification system were expanded to include the use of studies that used more than three animals in their evaluation.

All classification systems, as discussed previously, require the scoring of animals using the Draize scoring system (see **Table 4-1**). Scoring of animals occurs until the effect is cleared, but usually not beyond 21 days after the substance is applied to the eye of the animal. In order for a substance to be included in the evaluation of accuracy, fulfillment of three criteria was needed. These criteria were:

- At least three rabbits were tested in the study, unless a severe effect (e.g., corrosion of the cornea) was noted in a single animal. In such cases, substance classification could proceed based on the effects observed in less than three animals.
- A volume of 0.1 mL or 0.1 g was tested in each animal. A study in which a lower quantity was applied to the eye was accepted for substance classification, provided that a severe effect (e.g., corrosion of the cornea, lesion persistence) was observed in the animal.
- Observations of the eye must be made, at minimum, at 24-, 48-, and 72-hr following application of the test substance, if the lesion was not severe.

If any of the above three criteria were not fulfilled, then the data for that substance were not used for the accuracy analyses.

4.3.1 GHS Classification Rules Used for BRD Analysis

The classification of substances using the GHS classification system (UN 2003) proceeded in a stepwise fashion. Initially, each of the tested animals in the study was classified into one of four classification categories (Category 1, Category 2A, Category 2B, or nonirritant) based on their average corneal opacity, iritis, conjunctival redness and/or conjunctival chemosis score over Days 1, 2, and 3 after instillation of the substance. Once all animals were classified into one of these categories, the substance classification was determined based on the proportion of animals with a single classification.

For the animal classification, minimal Draize scores and time to clearing[2] of the effects used for this analysis are provided in **Table 4-3**. The criteria provided in this table are identical to those described in the GHS classification and labeling manual (UN 2003).

Table 4-3 Criteria Required for Classification of Animals into GHS Ocular Irritancy Categories (modified from UN 2003)

GHS Category	Animal Criteria Necessary for Classification
Category 1	Group A: - Effects in the cornea, iris, or conjunctiva that were not expected to reverse or did not fully reverse within the observation period of 21 days, or - A corneal opacity score of 4 at any time during the test Group B: - Average of the scores on day 1, 2, and 3 for opacity ≥ 3 and/or iris ≥ 1.5
Category 2A	- The average of the scores on day 1, 2, and 3 were: (a) 1 ≥ opacity < 3 or (b) 1 ≥ iritis < 1.5 or (c) conjunctival redness ≥ 2 or (d) chemosis ≥ 2 , and - The effect fully reversed within an observation period of normally 21 days
Category 2B	- The average of the scores on day 1, 2, and 3 was: (a) 1 ≥ opacity < 3 or (b) 1 ≥ iritis < 1.5 or (c) conjunctival redness ≥ 2 or (d) chemosis ≥ 2 , and - The effect fully reversed within 7 days
Nonirritant	The average of the scores on day 1, 2, and 3 for opacity and iris <1, and conjunctival redness and chemosis <2

Abbreviations: GHS = United Nations (UN) Globally Harmonized System.

After each animal tested for a substance was classified into one of the categories, the ocular irritancy potential of the substance was determined. As shown in **Table 4-4**, substance classification was dependent on the proportion of animals that produced the same response. As noted above, when necessary, the decision criteria for substance classification were expanded to allow for substances tested in more than three animals. In most of the cases, the proportionality needed to classify a substance was maintained (e.g., 1 out of 3 or 2 out 6

[2] For the purposes of this analysis, full reversal of the effects was defined as corneal and/or iritis scores of < 1 and redness and/or chemosis scores < 2. The rationale for this approach is that a substance is classified as a nonirritant unless the average corneal and/or iritis scores are at least 1 and the average redness and/or chemosis scores are at least 2.

animals were required for classification for most categories). However, in some cases, additional classification rules were necessary to include the available data. These additional rules are distinguished by italicized text in the table below.

Table 4-4 Criteria for Classification of Substances According to the GHS Classification System (modified from UN 2003)

GHS Category	Criteria Necessary for Substance Classification
Category 1	1. At least 1 of 3 animals or 2 of 6 animals classified as Category 1, Group A 2. *One of 6 animals classified as Category 1, Group A, and at least 1 of 6 animals classified as Category 1, Group B* 3. At least 2 of 3 animals or 4 of 6 animals classified as Category 1, Group B
Category 2A	At least 2 of 3 animals or 4 of 6 animals classified as Category 2A or Category 2B
Category 2B	At least 2 of 3 animals or 4 of 6 animals classified as Category 2B
Nonirritant	At least 2 of 3 animals or 4 of 6 animals classified as nonirritant

Abbreviations: GHS = United Nations (UN) Globally Harmonized System.

For substances where an unequivocal substance classification could not be made due to the response pattern of the tested animals (e.g., one animal classified as Category 1, Group B; two animals classified as Category 2B; three animals classified as nonirritant), the data were not used in the analysis.

4.3.2 EPA Classification Rules Used for BRD Analysis

The EPA classification system is dependent on the most severe incidence observed in an animal (EPA 1996). Unlike the GHS classification system (UN 2003), which focuses on the use of three or fewer animals, the EPA classification system requires at least three animals to be tested in most cases. Similar to the GHS classification system, substance classification proceeds in a stepwise manner. For the EPA classification system, a positive response is classified as an opacity or iritis score of equal to or greater than 1 or a redness or chemosis score of equal to or greater than 2. The observed score can occur at any time up to 21 days after substance application. Each tested animal is classified into one of four categories (Category I to IV). After each animal is classified, then substance classification followed. **Table 4-5** provides a listing of the criteria used for classification of animal responses into

Table 4-5 Criteria Required for Classification of Animals into EPA Ocular Irritancy Categories (EPA 1996)

EPA Category	Criteria for Animal Classification
Category I	- Corrosive, corneal involvement or irritation (iris or cornea score ≥ 1 or redness or chemosis ≥ 2) persisting more than 21 days or - Corneal effects that are not expected to reverse by 21 days
Category II	- Corneal involvement of irritation clearing[1] in 8-21 days
Category III	- Corneal involvement of irritation clearing in 7 days or less
Category IV	- Minimal or no effects clearing in less than 24 hours

Abbreviation: EPA = U.S. Environmental Protection Agency.

[1] For the purposes of this analysis, clearing was defined as iritis or cornea score < 1 and redness or chemosis score < 2.

one of the four ocular irritancy categories for this analysis. This classification system is identical to that described in the EPA labeling manual (EPA 1996).

Substance classification for the analysis described in this BRD was dependent upon the most severe classification observed. Thus, a single animal in the most severe category (Category I is more severe than Category II), led to classification of the substance to that category.

4.3.3 EU Classification Rules Used for BRD Analysis

The classification of substances using the EU classification system proceeded in a stepwise fashion (EU 2001). Similar to the GHS classification system (UN 2003), average Draize scores are used for classification purposes. However, compared to the GHS classification system, where the average for each animal is calculated separately, the calculation of average scores for the EU system was dependent upon the number of animals tested in the study. For those studies that used three animals, average corneal opacity, iritis, and conjunctival chemosis and/or redness scores were determined across the 24-, 48-, and 72-hour observation times for each animal tested. For those studies that used greater than three animals (i.e., six animals), the overall average opacity, iris, chemosis, and redness scores were determined across the 24-, 48-, and 72- hour observation times for all of the animals tested. Once these values were determined, the substance was classified based on the number of animals with a minimal positive average (for studies that used three animals) or the overall average (for

studies that used more than three animals). The criteria used for substance classification are provided in **Table 4-6**.

Table 4-6 Criteria for Classification of Substances According to the EU Classification System (EU 2001)

EU Category	Three Animals Tested	Greater than Three Animals Tested
R41	Two or more animals where the average animal Draize scores over Days 1, 2, and 3 were: Opacity ≥ 3 Iritis = 2 Or At least one animal (at end of observation period) where the effect has not reversed[1]	Overall mean animal Draize scores over Days 1, 2, and 3 were: Opacity ≥ 3 or Iritis > 1.5 Or At least one animal (at end of observation period) where the effect has not reversed
R36	Two or more animals where the average animal Draize scores over Days 1, 2, and 3 were: 2 ≤ Opacity < 3 1 ≤ Iritis < 2 Redness ≥ 2.5 Chemosis ≥ 2	Overall mean animal Draize scores over Days 1, 2, and 3 were: 2 ≤ Opacity < 3 1 ≤ Iritis < 2 Redness ≥ 2.5 Chemosis ≥ 2

Abbreviation: EU = European Union.

[1] For this analysis, the positive effect has not reversed when Opacity or Chemosis ≥ 2, Redness ≥ 2.5 or Iritis ≥ 1.

4.4 Availability of Original Records for the *In Vivo* Reference Data

Much of the published data on the prediction of ocular irritancy potential for test substances using the *in vivo* rabbit eye test method was limited to average score data (e.g., MAS, MMAS) or irritancy classification (e.g., mild, moderate, severe, or EU classification). An attempt was made to obtain the original records and/or compiled reports for the *in vivo* reference data. Although the original study records were not obtained for any of the studies, compiled *in vivo* data reports were obtained from the following organizations: 1) S.C. Johnson & Son, Inc. for the Swanson et al. (1995) and Swanson and Harbell (2000) studies; 2) the CTFA for the Gettings et al. (1996); 3) Access Business Group for the Casterton et al. (1996) study; and 4) ExxonMobil Biosciences, Inc. for the Bailey et al. (2004) study.

Additionally, individual animal data were available from the ECETOC eye irritation data bank (ECETOC 1998).

4.5 *In Vivo* Data Quality

Ideally, all data supporting the validity of a test method should be obtained and reported from studies conducted in accordance with GLP guidelines, which are nationally and internationally recognized rules designed to produce high-quality laboratory records (OECD 1998; EPA 2003a, 2003b; FDA 2003). These guidelines provide an internationally standardized approach for the conduct of studies, reporting requirements, archival of study data and records, and information about the test protocol, in order to ensure the integrity, reliability, and accountability of a study.

The extent to which the *in vivo* rabbit eye studies, used to provide the comparative data in the published BCOP validation studies, were compliant with GLP guidelines is based on the information provided in the published reports. Although an attempt was made to obtain the original study records, such records could not be obtained. Based on the available information, all of the reports appear to include *in vivo* data obtained according to GLP guidelines.

4.6 Availability and Use of Toxicity Information from the Species of Interest

Due to the possibility of irreversible eye injury that could impair vision or even result in blindness, human ocular irritancy studies are not routinely conducted. The only exceptions are for products that are intended for actual human eye use (e.g., contact lens solutions, ophthalmic pharmaceuticals) or cosmetic/personal care products that are known not to cause more than minimal to mild responses in animals. Bruner et al. (1998) and Cater et al. (2004) reported on studies of cosmetic and surfactant-based personal care formulations conducted in humans. However, all of the substances tested were classified as mild irritants or nonirritants. Procter & Gamble provided information from human exposures to three consumer product formulations as a comparison to the EU ocular toxicity classifications (EU

2001) that were assigned based on results from the low volume eye test (LVET) and the ICE test. However, because all three of these formulations were classified as nonirritants or mild irritants, based on results obtained in LVET and humans, evaluation of the accuracy of the BCOP test method for identifying ocular corrosives and severe irritants in humans is not possible.

It may be possible to consider accidental human exposure injury data to identify any substances or products that are capable of producing severe or irreversible eye injuries in humans. This data could then be compared with available animal testing data and hazard classifications to determine if the potential for severe human effects was not predicted by the animal test. A query to all ICCVAM regulatory agencies did not result in the identification of any substances or products that were know to produce a severe or irreversible human eye injury that was not predicted by the animal test. However, this lack of identified substances or products must be considered in light of weaknesses in the surveillance and reporting systems for such injuries. Several U.S. Federal agencies (OSHA, CPSC, and the National Institute for Occupational Safety and Health [NIOSH]) were contacted for data resulting from accidental human exposures. NIOSH estimated that there were approximately 39,200 chemical-related eye injuries in 1998, based on emergency department reports for work related eye-injuries (NIOSH 2004). Approximately 10,000 of these cases were attributed to an unidentified/unspecified chemical. Additional cases (<2500 each) were reported for injuries related to specific chemicals or chemical/product classes, which included[3]:

- acids (unspecified)
- adhesives/glues
- cement/mortar mix
- chlorine/chlorine bleach
- cleaning/polishing agents
- detergents/shampoos
- disinfectants
- drain/oven cleaners

[3] These specific chemicals or chemical/product classes are listed in alphabetic order; actual numbers of cases for each specific chemical or chemical/product class are not provided.

- gasoline/jet fuels/diesel fuel
- hydrochloric acid
- nonchlorine bleach
- paint removers/thinners
- paints
- soaps
- sodium hydroxide, potassium hydroxide, and potassium carbonate
- solvents/degreasers
- sulfuric acid

However, for the product classes listed above, specific information on which products were involved are not available. No human data were provided for any of these substances, nor were details of the types of ocular injuries sustained described.

In a study of adolescents (ages 6-17), Brevard et al. (2003) noted that of the 307 cases of acute occupational disinfectant-related illness reported to the Toxic Exposure Surveillance System and the California Department of Pesticide Regulation from 1993 to 1998, 51% (158/307) of the clinical manifestations involved the eye. Of these 158 cases, 150 included irritation/pain/conjunctivitis, 13 included blurred vision, 11 included corneal abrasions, 10 included corneal burns, and nine included tearing. The authors noted that halogens was the disinfectant class responsible for the majority of the overall cases (180/307), with hypochlorites accounting for 77% of these cases (139/180). However, it is not clear what proportion of these cases included ocular injuries. The authors concluded that working youths (ages 15-17) are four times more likely to suffer from occupational disinfectant-related illnesses than are adults (ages 25-44), suggesting a need for greater efforts to prevent such illnesses among youths.

4.7 Information About Accuracy and Reliability of the *In Vivo* Test Method

Given the lack of useful human ocular toxicity data, the accuracy of the *in vivo* rabbit eye test method for detecting ocular corrosives and severe irritants cannot be evaluated. However,

the reliability of the *in vivo* rabbit eye test in terms of the likelihood of classifying an ocular corrosive or severe irritant as a nonsevere irritant or nonirritant can be evaluated. Data for this analysis are being compiled from ocular corrosivity and irritation studies that followed the *in vivo* rabbit test method protocol described in **Section 4.1.1**. Data were received from several sources in response to personal communications with investigators, review of the Toxic Substances Control Act (TSCA) database, as well as in response to a published *Federal Register* Notice (Vol. 69. No. 57, pp. 13589-12861; published on March 24, 2004, available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm) requesting high quality *in vivo* rabbit eye test data. The results of this analysis will be made available when completed.

[This Page Intentionally Left Blank]